GWAS of variance QTL using monozygotic twins

# Analysis Plan

This document contains an analysis plan for a proposed collaborative project using genotyped monozygotic twins to identify genetic markers associated with differences in trait variance.

Given the wide range of proposed phenotypes and different research interests, generated summary data will contribute to two complementary but distinct manuscripts. One manuscript will focus on “physical” traits such as height, BMI and lipids (meta-analysis led by Kaprio and his team). The other manuscript will focus on psychiatric traits such as depression and ADHD (meta-analysis led by Assary).

If interested in contributing to the study, please contact the individuals listed below.

Github repository containing an example pipeline with relevant scripts can be found [here](https://github.com/LaurenceHowe/MZTwins-vQTL).

# Contact details

Questions about this analysis plan can be directed to:

Neil Davies; [neil.davies@bristol.ac.uk](mailto:neil.davies@bristol.ac.uk)

Laurence Howe; [laurence.howe@bristol.ac.uk](mailto:laurence.howe@bristol.ac.uk)

Jaakko Kaprio; [jaakko.kaprio@helsinki.fi](mailto:jaakko.kaprio@helsinki.fi)

Teemu Palviainen; [teemu.palviainen@helsinki.fi](mailto:teemu.palviainen@helsinki.fi)

Elham Assary; [e.assary@qmul.ac.uk](mailto:e.assary@qmul.ac.uk)

# Background and research question

The majority of genome-wide association studies (GWAS) to date have focused on identifying genetic markers associated with phenotype means. Contrastingly, much less is understood about genetic markers which influence phenotypic variation (variance quantitative trait loci; vQTL). The association of such markers with phenotypic variation may reflect interactions with environmental factors, as is thought to be the case for the adiposity influencing variant near/in *FTO*, or may capture biological mechanisms of an allele leading directly to higher variability. Therefore, identification of vQTL could improve understanding of underlying biology as well as the interplay between genetics and the environment.

Previous research into vQTLs has largely focused on obesity-related traits and predominantly used non-family based methods; for example, two recent studies (Young, Wauthier et al. 2018, Wang, Zhang et al. 2019) using UK Biobank data. An orthogonal approach to identify vQTL is to use a family-based method where we evaluate if there are genetic effects on phenotypic differences between monozygotic (MZ) twins. MZ twins share the same genotype at the vast majority of the genome but will often have discernible phenotypic differences. This MZ twin approach for identifying vQTL has been used previously for lipids on a sample of around 2,000 twin-pairs (Surakka, Whitfield et al. 2012) and in emotional symptoms in a sample of around 1,000 twin-pairs (Keers, Coleman et al. 2016). In this more recent study, the authors showed that a polygenic score created using the results of the GWAS moderated the effects of parenting and treatment response in two further samples.

We propose to extend previous work with a larger sample size and with the inclusion of additional phenotypes. This will involve a collaborative meta-analysis GWAS of associations between genetic variation and phenotypic differences between monozygotic twins for a variety of physical and psychological phenotypes (e.g. height, body mass index, lipids, depression and anxiety and educational attainment).

# Phenotypes and definitions

## List of phenotypes

Below is the list of phenotypes that we propose to include in analyses, with suggested phenotype definitions provided in Appendix 1.

1. Height
2. Body mass index
3. Educational attainment
4. Systolic blood pressure
5. Lipids
6. LDL-C
7. HDL-C
8. Triglycerides
9. Waist-hip-ratio, adjusted for BMI
10. Performance on cognitive tests
11. Alcohol consumption
12. Depressive symptoms
13. Anxiety symptoms
14. Neuroticism
15. Subjective well-being
16. ADHD symptoms
17. Autism symptoms
18. Psychotic -like experiences

We omitted phenotypes which were non-continuous or which we believed would have low sample sizes (e.g. cigarettes per day). Please contact us if you would like to propose additional phenotypes to include in analyses.

## Phenotype coding

The nature of the phenotype variation focused project requires continuous or categorical non-binary phenotypes. In general, we would advocate using continuous phenotypes where possible and recommend inverse normal transformation and standardisation of the absolute phenotypic difference (APD) that is corrected for age, sex and 10 principal components. This is because absolute phenotypic difference (APD) data are highly skewed, and the range differs according to the various measures used across studies.

We leave it to the discretion of the individual cohorts regarding selecting the most appropriate variable when there are differences in degree of continuity and sample sizes between measures for the same phenotype. Please get in touch with the individuals listed as contacts above, if you need help with selecting the most suitable measures. In longitudinal studies, phenotypes may have been measured more than once. Where data are available for study participants in more than one of the three categories (childhood, adolescence and adulthood: see definitions below) we propose that analyses should be run separately in available categories using timepoint with the most data. Multiple analyses in the same individuals may be informative about age-related heterogeneity.

Imputation of missing data for incomplete pairs is not recommended. Please use the timepoint with largest complete twin-pairs instead.

1) Childhood (aged 5-12 year)

2) Adolescence (aged 13-18 years)

3) Adulthood (aged over 18 years)

**4.3. Covariates:**

The covariates include the first 10 Principal Components of the sample, age and sex. The MZ difference score is adjusted for these covariates prior to rank transformation. Any additional covariates, such as chip or batch effects can be added at this stage, along with age, sex and PCs.

Please note age should reflect the age of the twins at the time of data collection/response rather than their current age.

# Consortium opt-in procedure

To participate in the study, please contact the individuals listed as contacts above. See below for information on eligibility criteria. If you have any questions about phenotypes, please contact us.

# Eligibility criteria

## Study

The study must include at least 100 pairs of MZ twins satisfying ALL of the following criteria.

1. one or both twins must be genotyped.
2. imputed genotype data is available (e.g. 1000 Genomes or HRC).
3. both twins must have complete data for one or more phenotypes.
4. both twins must have complete covariate data (age, sex, and principal components for the genotyped twin).
5. samples are of European ancestry.

## Phenotype inclusion

Phenotypes should only be included if there is complete data for at least 100 MZ twin pairs.

# Genotypes & imputation

Studies must have genotype data from all 22 autosomes imputed to either the 1000 genomes reference panel (preferably phase 3) or Haplotype Reference Consortium (HRC). Where imputation to both has been completed, our preference is for HRC over 1000 genomes. If your data has been imputed to a reference panel other than HRC or 1000 genomes, please contact us to discuss this. We suggest quality control on imputed SNPs relating to imputation quality (eg, *r*² and INFO metrics of MACH and IMPUTE, respectively) > 0.3 for hapmap imputed data and > 0.5 for 1000G or HRC data. We also recommend filtering markers on call rate (> 95%) and allele frequency (MAF > 1%).

We suggest performing QC on relatedness (between different MZ pairs rather than within the twin pairs). Please remove one pair at random when there are two MZ pairs with kinship > 0.1.

# Models used to test for association

The underlying model involves a regression of the absolute phenotypic difference (APD) (e.g. in R: “abs(Height1-Height2)”) between MZ twins on the genetic marker. The APD is residualised for sex, age and 10 principal components, then residuals are inverse rank transformed and standardised. This transformed variable is used as the predictor in the following models:

a) Model 1 (primary analysis): transformed APD

b) Model 2: Model 1 and control for within-pair mean (WPM)

c) Model 3: Model 1 without residualizing on principal components.

d) Model 4: Model 1 without sex adjustment, stratified by sex.

*APD\_nx ~ G*

PLINK v 1.9 will be used to run all GWAS using best guess genotype data and an additive model.

# Data exchange Procedure

We will provide scripts to upload the GWA results to a secure server (TBD). We will provide studies with bespoke access links once results are ready for submission.

The sharing of individual-level data is not necessary as meta-analyses will be performed using summary data.

## Results to return

The GWAS summary statistics file should contain the information listed in the table below. Please also complete the excel sheet found here and submit along with results to the repository.

|  |  |
| --- | --- |
| Variable name (case sensitive) | Description |
| CHR | Chromosome. |
| BP | Base Pair. |
| A1 | The effect allele. For example, for an A/G SNP in which AA=0, AG=1 and GG=2, the effect allele is G. |
| A2 | The other allele. For example, for an A/G SNP in which AA=0, AG=1 and GG=2, the other allele is A. |
| MAF | The allele frequency of the minor allele. |
|  |  |
| Beta | Beta estimate from genotype-phenotype association. |
| Se | Standard error estimate from genotype-phenotype association. |
| Pval  N  INFO | P value of test statistic from genotype-phenotype association.  Sample size for each variant  Measure of imputation quality |

# Meta-analysis

Once GWAS have been performed and results returned, we will check results for potential errors before conducting GWAS meta-analyses using a random-effects model on betas and standard errors in appropriate software, likely METAL.

We will evaluate potential heterogeneity (e.g. between studies or by age group), i.e. using Cochran’s Q.

## Primary analysis

The primary analysis will be to identify loci associated with phenotypic variation across the different phenotypes. Previous research has found evidence of gene-environment interactions involving BMI loci such as *FTO* so we will use BMI as a positive control.

## Secondary analyses

Replication of previously published findings (e.g. BMI).

Heritability analyses such as LDSC: What percentage of the phenotypic variation between MZ twins is explained by genotype?

Genetic correlation using LDSC: What is the genetic correlation with other phenotypes?

Functional lookups.

Testing key loci for evidence of gene-environment interactions in UK Biobank.

Creating polygenic scores for sensitivity to be tested in GxE models in further samples (e.g. UK Biobank)

# Significance

For determining loci of interest, we will use the conventional Genome-wide significance threshold of 5x10-8. We will evaluate if key loci are detected using complementary vQTL approaches in non-related individuals. For the secondary analyses, we will use appropriate corrections such as Bonferroni corrected p-values based on the number of phenotypes tested for LDSC analyses.

# Archiving and deposition of results

We will deposit full results from the GWAS, from the subset of studies that allow publication of summary data, in the GWAS catalogue for release on publication for use in downstream analysis.

# Manuscript authorship

For each paper, PIs determine the authors from their group using the widely accepted criteria for authorship on scientific papers ([icmje criteria](http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html)).

Each study contributing GWA data can name authors. Authors will be ordered based on contributions, which will be determined by assigning authors to each of the following tiers:

1. Central analysis and writing group (junior).
2. Study-specific lead analyst.
3. Study-specific sample collection and data generation (phenotyping/genotyping/supporting analysis).
4. Study PIs + study design.
5. Central analysis and writing group (senior).

The author list concludes with “the within family consortium”. This term has no attached author or affiliation. The studies will be ordered according to their sample size, from largest to smallest (tier II) or smallest to largest (tiers IV, V). Where there are multiple authors from one study in the same tier, authors are put in alphabetical order.

We plan to write two papers. One on neuropsychiatric phenotypes, one on the other phenotypes. Elham Assary and Patricia Munroe will lead on the neuropsychiatric paper, Jaakko Kaprio and Teemu Palviainen will lead on the other phenotypes.

# References

Keers, R., J. R. Coleman, K. J. Lester, S. Roberts, G. Breen, M. Thastum, S. Bögels, S. Schneider, E. Heiervang, R. Meiser-Stedman, M. Nauta, C. Creswell, K. Thirlwall, R. M. Rapee, J. L. Hudson, C. Lewis, R. Plomin and T. C. Eley (2016). "A Genome-Wide Test of the Differential Susceptibility Hypothesis Reveals a Genetic Predictor of Differential Response to Psychological Treatments for Child Anxiety Disorders." Psychother Psychosom **85**(3): 146-158.

Surakka, I., J. B. Whitfield, M. Perola, P. M. Visscher, G. W. Montgomery, M. Falchi, G. Willemsen, E. J. de Geus, P. K. Magnusson, K. Christensen, T. I. Sørensen, K. H. Pietiläinen, T. Rantanen, K. Silander, E. Widén, J. Muilu, I. Rahman, U. Liljedahl, A. C. Syvänen, A. Palotie, J. Kaprio, K. O. Kyvik, N. L. Pedersen, D. I. Boomsma, T. Spector, N. G. Martin, S. Ripatti and L. Peltonen (2012). "A genome-wide association study of monozygotic twin-pairs suggests a locus related to variability of serum high-density lipoprotein cholesterol." Twin Res Hum Genet **15**(6): 691-699.

Wang, H., F. Zhang, J. Zeng, Y. Wu, K. E. Kemper, A. Xue, M. Zhang, J. E. Powell, M. E. Goddard, N. R. Wray, P. M. Visscher, A. F. McRae and J. Yang (2019). "Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank." 519538.

Young, A. I., F. L. Wauthier and P. Donnelly (2018). "Identifying loci affecting trait variability and detecting interactions in genome-wide association studies." Nat Genet **50**(11): 1608-1614.

# Appendix 1: Phenotype definitions

Please note that in the first instance, these phenotype definitions are taken from the separate within-families sibling GWAS project so definitions and leads may be subject to change.

### 1.   Height

Lead: Laurence Howe

The participants height while standing measured in centimetres. If height has been recorded with greater precision, round to the nearest centimetre. Where height data have been collected from multiple sources, our preference is that measures from direct assessments of cohort participants are used above self-reported height.

### 2.   BMI

Lead: Laurence Howe

Please calculate BMI as weight in kilograms divided by standing height in metres squared (BMI = kg/m2). Where weight has been collected with greater precision than kilograms, please round to the nearest kilogram. Where height has been collected with greater precision than centimetres (0.01 metre), please round to the nearest centimetre. Where weight and height data have been collected from multiple sources, our preference is that measures from direct assessments of cohort participants are used above self-reported measures.

### 3.   Education

Lead: Tim Morris

For years of education, completed academic qualifications should be mapped to levels of the International Standard Classification of Education (ISCED) and converted to the corresponding completed years of education required for completion of the qualification in US years of schooling (Table 1). If ISCED levels 5 and 6 cannot be distinguished, please set values for those with tertiary education as 20 years of schooling.

*Table 1: ISCED code mapping to US years of schooling.*

|  |  |  |
| --- | --- | --- |
| Qualification | ISCED Code | US years of schooling |
| Pre-primary education | 0 | 1 |
| Primary education or first stage of basic education | 1 | 7 |
| Lower secondary or second stage of basic education | 2 | 10 |
| Upper secondary education | 3 | 13 |
| Post-secondary non-tertiary education | 4 | 15 |
| First stage tertiary education (not leading to advanced research qualifications) | 5 | 19 |
| First or second stage tertiary education (cannot distinguish first from second) | 5 | 20 |
| Second stage tertiary education (leading to advanced research qualifications) | 6 | 22 |

### 4.   Systolic blood pressure

Lead: Neil Davies

The participants systolic blood pressure adjusted for treatment with anti-hypertensives. Please use the time point that has the least missing data across your sample. Please replace missing values with values from other measurement occasion if multiple time points are available. If multiple measurements of blood pressure were taken at each clinic visit, please use the average of the measurements. If some participants have been treated with treated with anti-hypertensive medication, please add 10mmHg to the blood pressure of individuals who are on treatment at the time of measurement.

### 5. Lipids

Lead: Ben Brumpton

LDL: The participants untreated LDL, HDL and triglyceride levels. Please replace missing values with values from other measurement occasion if multiple time points are available. If multiple measurements of cholesterol were taken at each clinic visit, please use the average of the measurements. Please normalise to mean zero standard deviation one.

### 6.   Waist-hip ratio, adjusted for BMI

Lead: Neil Davies

Please calculate waist-hip ratio adjusted for BMI by residualizing using linear regression. Waist and hip measures should be rounded to the nearest centimetre. Where multiple waist and hip measures are available our preference is for direct or clinic measurements above self-report. Please ensure that the waist-hip measurements and BMI were made concurrently. If there are multiple measurements over time, please choose the measurement occasion with the least missing data across your cohort.

### 7.   Performance on cognitive tests

Lead: Laurence Howe / David Hill

A general cognitive function score will be created, where appropriate data are available in the sample, by Principal Component Analysis (PCA) using at least 3 cognitive tests that assess different cognitive domains. Only one score should be used from each cognitive test. The tests should not include the clinical cognitive assessments that are used as screening instruments for dementia (e.g., MMSE). The score to be created and used for the genetic analysis will be the first unrotated principal component. We have supplied a ‘boilerplate text’ paragraph so that samples’ analysts can fill in certain key outcomes of their PCA. The general cognitive function score from the PCA (first unrotated principal component) should be saved as a standardized variable (mean = 0; SD = 1). The variable must be in the direction of higher positive scores indicating better general cognitive performance.

### 8.   Alcohol consumption

Lead: Laurence Howe

The average number of alcoholic units consumed per week as reported by the participant, aggregated across all types of alcohol. If numbers of drinks have been recorded as categorical (e.g. 1-5 drinks per week), set value as the mid-point of the range (e.g. 2.5). Please remove extreme outliers (e.g. 5 S.D. greater than the mean).

### 9. Depression – depressive symptoms

Lead: Elham Assary

Depression is a highly heterogeneous disorder with many clinical presentations. The diagnosis requires a distinct change of mood, characterized by sadness or irritability, accompanied by psychophysiological changes, such as disturbances in sleep, appetite, or sexual desire; constipation; loss of the ability to experience pleasure in work or with friends; crying; suicidal thoughts; and slowing of speech and action.

A great variety of rating scales have been developed to assess symptoms of depression. The scales differ in the number of items included and with respect to the types of symptoms assessed: mood symptoms (depressed mood and irritability), behavioral symptoms (suicide and anhedonia), somatic symptoms (appetite disturbance, sleep disturbance, low energy, and psychomotor retardation or agitation), cognitive symptoms (hopelessness and worthlessness), and concentration symptoms (poor concentration and decision-making). Frequently used scales include: the Hamilton Rating Scale for Depression (HRSD), the Beck Depression Inventory (BDI), the Center for Epidemiologic Studies Depression Scales (CES-D), the Quick Inventory of Depressive Symptoms (QIDS), and the Self-Rating Depression Scale (SDS) and Hospital Anxiety and Depression Scale (HADS). Depressive symptoms in children are commonly measured by the strengths and difficulties questionnaire (SDQ) in children, and the short mood and feelings questionnaire (SMFQ), and adolescents with SMFQ and adult behaviour checklist (ABCL).

If multiple rating scales of depressive symptoms are available, select the scale with most items (tapping most symptom domains). Please sum the items and treat the final summed variable as a continuous measure of depression symptoms. Please ensure that measures are coded so that higher values represent higher symptom levels.

### 10. Anxiety symptoms

Lead: Elham Assary

Anxiety is a hetroegentous disorder, with clinical diagnoses consisting of specific anxiety disorders (e.g. phobias, post-traumatic stress disorder, social anxiety disorder) and generalised anxiety disorder (GAD). We are interested in generalised anxiety symptoms, usually measured via self- or other reports. There are a wide range of screening measures available for different age groups: for adults, Generalised Anxiety Disorder scale (GAD) and anxiety subscale of the HADS; for adolescents, ABCL anxiety subscale; for children, anxiety subscale of the SDQ and Screen for Child Anxiety Related Emotional Disorders. Please ensure that measures are coded so that higher values represent higher symptom levels.

### 11. Neuroticism

Lead: Elham Assary

Neuroticism is a fundamental domain of personality functioning and structure. The personality trait refers to a lack of emotional stability; stress vulnerability; the tendency to experience intense negative emotions, affects, and cognitions; and impulsive behaviors under emotional strain. Neuroticism can be validly measured from age 6-7 years. The Hierarchical Personality Inventory for Children is one of few instruments specifically designed to assess neuroticism in children. For adults, scales include the NEO-Personality Inventory, the Eysenck Personality Questionnaire, and various scales using Big-Five factor markers from the International Personality Item Pool (IPIP- including the 300-item representation of the Revised NEO Personality Inventory). Most modern neuroticism scales show strong overlap in item content (although they consist of different items), and the sum scores correlate highly across inventories. Scales usually agree on lower-order traits (e.g. anxiety-withdrawal, depression-unhappiness, vulnerability-stress reaction). However, there is less agreement whether aggression, impulsivity, inferiority, and dependency belong to the neuroticism domain.

 Please sum the neuroticism items and treat the final summed variable as a continuous measure. Please ensure that higher values on the scale represent higher levels of neuroticism.

### 12. Subjective wellbeing

Leads: Elham Assary

Reporting of subjective wellbeing is likely to be highly variable across studies. Our preference is for subjective wellbeing to have been measured using a battery of questions. If your study has done this, please sum responses and treat this final summed variable as a continuous measure of subjective wellbeing. If data on subjective wellbeing are only available from a single question reported on a Likert scale, please treat the response variable as continuous. Please ensure that measures are coded so that higher values represent higher levels of subjective wellbeing.

13. ADHD symptoms

Lead: Elham Assary

Attention-deficit hyperactivity disorder (ADHD) is defined in DSM-5 and ICD-11 as a childhood-onset neurodevelopmental disorder of attention, activity and impulsivity. ADHD symptoms commonly persist into adulthood.

ADHD symptoms are typically measured continuously based on DSM-based rating scales, often with separate scales for attention problems and Hyperactivity/impulsivity which can be summed into combined ADHD symptoms. Some rating scales include items corresponding to the full DSM-IV/DSM-5 criteria for ADHD, for example the Conners’ Rating Scale or the Rating Scale for Disruptive Behavior Disorders (RS-DBD). Other rating scales contain fewer items related to ADHD, such as the the Attention Deficit/Hyperactivity Problems scale of the Child Behavior Checklist (CBCL) or hyperactivity-inattention scale of the Strengths and Difficulties Questionnaire (SDQ). Our largest sample, the Swedish twins, use Autism - Tics, AD/HD and other Comorbidities inventory (ATAC) in children. SDQ in adolescents and Adult ADHD Self-Report Scale (ASRS) in adults. To ensure consistency, the preference is to use the data from these scales, if you have data from multiple instruments. Please ensure that measures are coded so that higher values represent higher symptom levels.

The genetic influences on ADHD symptoms may vary with age and previous GWAS have restricted their samples to specific age groups (e.g., children younger than 13 years [PMID: 27663945]). Therefore, separate analyses will be performed for ADHD symptoms in childhood (<13 years), adolescence (14-17 years) and adults (18 years and older).

For each of the 3 age groups (sequential criteria):

·       If multiple rating scales of ADHD symptoms are available, select the most comprehensive scale (most items related to ADHD).

·       If multiple informant reports are available, prefer caregiver report > teacher report > self-report for consistency and due to lower reliability and heritability of self-report in children and adolescents (PMID: 29892054).

·       If repeated measures are available within the age group, select the measure closest to age 9 in school-aged children, age 15 in adolescents. For adults, early adulthood (age 20) and later (age >30) in adults.

14. Autism symptoms

# Autism spectrum disorders are neurodevelopmental disorders broadly reflecting difficulties in social interaction and verbal communication and repetitive behaviours. Symptoms emerge in early chilhood, but diagnosis is often delayed. Different screening questionnaires are developed which distinguish between the sub-categories (e.g. Asperger, developmental delay non-specified). For the current study, we are interested in Autism symptoms specifically, commonly assessed via measures such as Autism Quotient (AQ), Childhood Autism Rating Scale (CARS) and ATAC. Please ensure that measures are coded so that higher values represent higher symptom levels.

15. Psychotic -like experiences (PLE)

While the worldwide prevalence of psychosis and schizophrenia disorders is low (1-4%), psychotic-like experiences are prevalent in community and non-clinical samples (~20%-30%). PLEs include sub-clinical threshold, persecutory ideation or perceptual abnormalities, that occur in community populations. PLEs are associated with several other psychiatric disorders including mood and anxiety disorders. PLEs are screened using self/other report questionnaires, or interviews, that cover some or all of these domains: paranoia, hallucinations, cognitive disorganization, anhedonia, negative symptoms. Measures such as the Community Assessment of Psychic Experiences (CAPE-42) are used with adults, whereas Kiddie - Schedule for Affective Disorders and Schizophrenia for school-aged children is used with children and adolescents. Please ensure that measures are coded so that higher values represent higher symptom levels.

# Appendix 2: Contacted studies

The following studies have been contacted previously, let us know if there are other studies we should contact:

1. Finnish Twin Cohort
2. Swedish Twin Registry
3. Texas Twin Project
4. QIMR
5. Murcia Twin Registry
6. NTR
7. Australian Mammographic Density Twins and Sisters Study
8. Italian Twin Registry
9. Minnesota Center for Twin and Family Research
10. Osaka University Twin Registry
11. Longitudinal Study of Aging Danish Twins
12. TwinsUK
13. TEDS
14. Danish Twin Registry (DTR)
15. QNTS
16. Mid-Atlantic twin reg
17. TwinGene
18. The Healthy Twin Study (Korea)
19. Korean Twin Study
20. Chinese National Twin Registry
21. Colorado Twin Registry
22. UK Biobank
23. MoBa
24. CKB
25. AddHealth